

АНАЛИЗ ЭМОЦИЙ ЧЕЛОВЕКА ПО ПОРТРЕТНОМУ СНИМКУ

Совпель Д. С.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Калугина М. А. – канд. физ.-мат. наук, доцент

Рассмотрены алгоритмы построения моделей, основанных на применении сверточных нейронных систем, для классификации эмоций портретного снимка человека. Описана работа моделей, условие их применения в зависимости от выбранного подхода к построению алгоритма, представлены результаты и возможные перспективы улучшения модели.

Многие задачи искусственного интеллекта и машинного обучения сводятся к автоматизации обработки большого количества данных. Одной из таких проблем является определение эмоций на основе его снимка или видеофрагмента. Такая система имеет множество сфер применения: система наблюдения за озлобленными гражданами позволит предотвратить возможное преступление, HR в компаниях смогут определить незаинтересованных кандидатов на работу. Не всегда возможно выполнять такие задачи “вручную”, так как объем данных для обработки увеличивается, и скорость классификации эмоций человеком несоизмерима с количеством данных.

В основе всех алгоритмов была выбрана классическая структура построения моделей обработки изображений – использование сверточных нейронных сетей. Для таких задач возможен выбор нескольких подходов: использование уже тренированной архитектуры и разработка и обучение собственной архитектуры. Поэтому было выбрано два основных алгоритма: модель на базе архитектуры ResNet и собственная модель. Для построения модели был использован датасет FER 2013 [1]. Он состоит из данных, разбитых на множества для тренировки и валидации (теста) модели. Каждая картинка представляет собой черно-белое изображение 48 на 48 пикселей. Все изображения разделены на 7 классов: злорада, отвращение, испуг, счастье, грусть, удивление и спокойствие. Для того чтобы использовать такие изображения для обучения, было увеличено количество каналов (так как сверточные сети используют цветные изображения с 3 цветовыми каналами) путем копирования изображения 3 раза.

Модель на базе ResNet [2] включает в себя дообученные слои модели ResNet и слоев для дообучения сети до конкретной задачи классификации. Такой подход позволяет, во-первых, использовать уже предтренированные слои для выделения общих компонент изображения, так как первые слои данной модели выполняют именно эту задачу, во-вторых, уменьшить вычислительную сложность для тренировки модели под конкретную задачу при помощи добавленных слоев. Поэтому к архитектуре ResNet был добавлен один слой субдискретизации и 3 полносвязных слоя с регуляризацией дроп-аутом. Также часть последних слоев исходной архитектуры была также использована для дообучения под классификацию эмоций.

Вторая модель построена следующим образом: три слоя, состоящие из сверточного слоя нормализации по батчу, функции активации и слоя максимальной субдискретизации, а затем 3 полносвязных слоя с дроп-аутом для регуляризации. Между слоями с параметрами для обучения расположены слои для регуляризации. Это позволяет предотвратить переобучение модели на относительно маленьком датасете.

Для реализации моделей использовался язык Python 3.8, библиотека Pandas для обработки датасета FER 2013 и библиотека для построения глубоких сетей Keras. Эти инструменты позволяют наиболее быстро реализовывать, тренировать и оценивать модели нейронных сетей. Модель с данной архитектурой показала точность в 57% на валидационном множестве (рис.1).

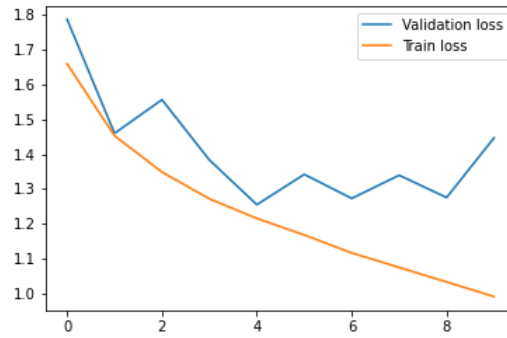


Рисунок 1 - график скорости обучения модели с разработанной архитектурой

Модель с архитектурой на базе ResNet показала точность в 68% на валидационном множестве (рис.2).

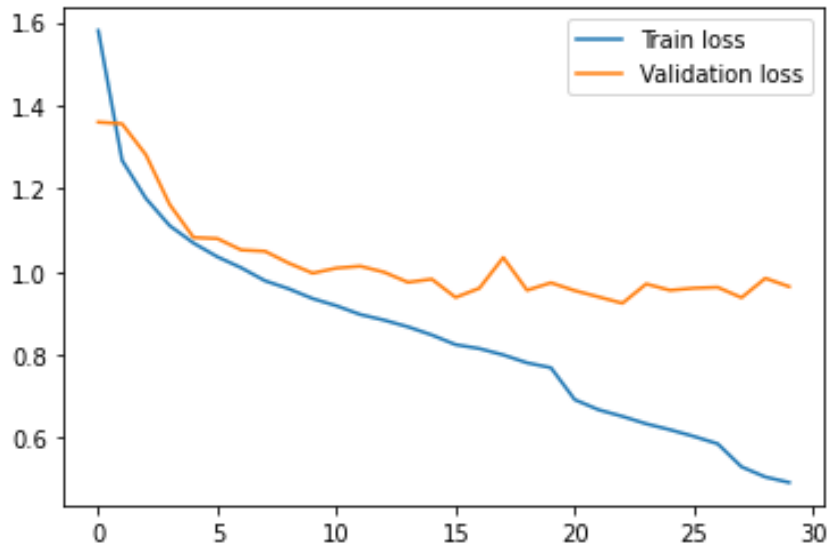


Рисунок 2 - график скорости обучения модели с архитектурой на основе ResNet



```
(dl_py38) $ python predict.py -i angry_face.jpg -m resnet18
angry: 99.36%
disgust: 0.00%
fear: 0.26%
happy: 0.12%
sad: 0.26%
surprise: 0.00%
neutral: 0.00%
```

Рисунок 3 - пример работы алгоритма на основе ResNet

Таким образом, построенные модели могут использоваться для классификации снимков человека. Данные модели можно использовать в связке с алгоритмами обнаружения лица, которые позволяют выделить лицо человека, для систем непрерывной видеосъемки, усредняя множество кадров и получая одно изображение для классификации (рис.3). В результате можно построить систему непрерывного наблюдения за эмоциями человека.

Список использованных источников:

1. Dataset fer2013 // [Электронный ресурс] <https://www.kaggle.com/ahmedmoorsy/facial-expression>
2. Deep Residual Learning for Image Recognition / K. He [et al.] // arXiv:1512.03385v1 [cs.CV], 2015.